Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization

Dutilh Novaes, Catarina; Reck, Erich

Published in:
Synthese

DOI:
10.1007/s11229-015-0816-z

Publication date:
2017

Synthese (2017) 194:195–215
DOI 10.1007/s11229-015-0816-z

S.I. : CARNAP ON LOGIC

Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization

Catarina Dutilh Novaes[1] · **Erich Reck**[2]

Received: 15 August 2014 / Accepted: 25 June 2015 / Published online: 10 July 2015

Abstract *Explication* is the conceptual cornerstone of Carnap's approach to the methodology of scientific analysis. From a philosophical point of view, it gives rise to a number of questions that need to be addressed, but which do not seem to have been fully addressed by Carnap himself. This paper reconsiders Carnapian explication by comparing it to a different approach: the 'formalisms as cognitive tools' conception (Formal languages in logic. Cambridge University Press, Cambridge 2012a). The comparison allows us to discuss a number of aspects of the Carnapian methodology, as well as issues pertaining to formalization in general. We start by introducing Carnap's conception of explication, arguing that there is a tension between his proposed criteria of fruitfulness and similarity; we also argue that his further desideratum of exactness is less crucial than might appear at first. We then bring in the general idea of formalisms as cognitive tools, mainly by discussing the reliability of so-called statistical prediction rules (SPRs), i.e. simple algorithms used to make predictions across a range of areas. SPRs allow for a concrete instantiation of Carnap's fruitfulness desideratum, which is arguably the most important desideratum for him. Finally, we elaborate on what we call the 'paradox of adequate formalization', which for the Carnapian corresponds to the tension between similarity and fruitfulness. We conclude by noting that formalization is an inherently paradoxical enterprise in general, but one worth engaging in given the 'cognitive boost' it affords as a tool for discovery.

✉ Catarina Dutilh Novaes
c.dutilh.novaes@rug.nl

Erich Reck
erich.reck@ucr.edu

1 Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands

2 Department of Philosophy, University of California, Riverside, Riverside, CA 92521, USA

Keywords Carnap · Explication · Statistical Prediction Rules · Formalization · Cognitive artifacts · Enlightenment

The first half of the twentieth century was a period of great enthusiasm for the recently developed methods of formal logic and formal axiomatics (Awodey and Reck 2002). Pioneers in this tradition—such as Russell, Hilbert, Zermelo, Tarski, Church, Carnap, Quine, and earlier Frege, Peano, and Dedekind, among others—were busy engaging 'hands-on' with projects of formalization concerning specific concepts/topics/areas, in philosophy, mathematics, and beyond. Perhaps because this was a burgeoning field, they were also led to reflect on basic methodological questions, including the scope and limits of formal methods, criteria of adequacy for formalization, and so forth. Such meta-theoretic reflections have continued until today, even though they are often kept to a minimum.

Rudolf Carnap is undoubtedly one of the main figures in this tradition. He too spent relatively little time on methodological discussions, preferring to put formal methods to use rather than talking about them. However, in Chap. 1 of his book *Logical foundations of probability* (1950), he offers a seminal discussion of the uses and methodology of formal methods, under the guise of the notion of *explication*. To be sure, Carnapian explication does not pertain exclusively to formalization; it covers the process of explicating vague, informal concepts through more precise, scientifically informed accounts more generally. But Carnap seemed to view formalization by logical means as the quintessential form of explication, especially given his focus on *exactness*; this view then became very influential.

Today, more than 60 years later, explication still represents an important approach to formalization in its benefits and drawbacks. It could well, and sometimes explicitly does, serve as the methodological basis for current investigations utilizing formal tools, again in philosophy and elsewhere (Kuipers 2007; Maher 2007; Justus 2012). From a philosophical point of view, explication also gives rise to a number of questions that need to be addressed but were treated insufficiently by Carnap himself. In particular, what is arguably the key criterion of adequacy for formalization on the explication model, namely fruitfulness, is somewhat under-developed by Carnap, including its relationship to the other criteria.

In what follows, we will reconsider Carnapian explication by comparing it to a second approach to formalization: the 'formalisms as cognitive tools' conception recently defended in detail by one of the two present authors (Dutilh Novaes 2012a). The latter is a true heir to the Carnapian tradition, even though there are also important differences. A comparison between the two is meant to clarify a number of aspects of the Carnapian methodology—fruitfulness in particular. The 'formalisms as cognitive tools' conception of formalization might even be viewed as an *explication* of Carnapian explication itself, in virtue of grounding Carnap's ideas in a wealth of empirical results pertaining to human cognition. Yet the comparison will also bring out a problem, or paradox, inherent to both approaches.

To prevent a possible misunderstanding right away, let us point out that we do not claim to present a new textual exegesis of Carnap's work. Indeed, our goal is not primarily that of textual analysis, though at different times we also raise (and partially address) a few interpretative issues as well. As far as textual exegesis goes, we rely

extensively on the interpretation given in Carus (2007), even though we know that it is not unanimously accepted among Carnap scholars. But we find Carus' interpretation to be historically well grounded and philosophically compelling, as well as eminently suitable for our purposes. Indeed, it offers an appropriate vantage point for the general discussion of the notion of formalization that we propose. In effect, in this paper we take Carnapian explication as a starting point to discuss the concept of formalization as such, in some of its salient aspects.[1] Its goal is thus mainly systematic and philosophical, rather than interpretive and historical.

More specifically, we proceed by adopting an *empirically informed* approach, which may or may not be faithful to the historical Carnap.[2] We in fact believe that such an approach is very much in the spirit of Carnap's own conception of explication, i.e. to formulate an account of a given concept that is more precise and more scientifically grounded than the initial, informal concept. But since our primary goal is not to defend a textual exegesis of Carnap's thought, even those interpreters who do not endorse our particular reading of Carnap's enterprise may find our analysis illuminating in some respects (or so we hope), provided they bear in mind its non-exegetical chief purpose.

The paper proceeds as follows: In the first section, we present the basics of Carnap's conception of explication, starting with a brief comparison between his approach and Tarski's. We then argue that there is a tension between the criteria of fruitfulness and similarity adopted by Carnap for explication, and also that his further desideratum of exactness is less fundamental than might appear at first. In the second section, we bring in the general idea of formalisms as cognitive tools. We do so mainly by discussing the reliability of so-called statistical prediction rules (SPRs), i.e. simple algorithms used to make predictions across a range of areas, as an illustrative example of formalization. In the third section, we elaborate on what we call the 'paradox of adequate formalization', which for the Carnapian corresponds to the tension between similarity and fruitfulness. We conclude by noting that, while formalization is an inherently paradoxical enterprise in general, it is still very much worth engaging in given the 'cognitive boost' it affords as a tool for discovery.

1 Carnapian explication

1.1 Tarski on criteria of adequacy for formalization

As already mentioned, some of the pioneers in applying formal/mathematical methods to philosophical problems did pay at least some attention to the issue of what counts as an adequate formalization. Alfred Tarski, for example, sought to tie his formal accounts to the philosophical, informal concepts they were supposed to be accounts *of* by means of the concept of 'conditions of material adequacy'. It will prove instructive to briefly examine Tarski's ideas in this respect before moving on to Carnap, as the

[1] Even though explication and formalization are not identical, as we acknowledge along the way, we move back and forth freely between the two when the argument and the context warrant it.

[2] Such approaches are often described as 'naturalistic', but we consider this term to have ambiguous connotations as well as problematic undertones. We thus prefer 'empirically informed'.

resulting contrast between these two pioneers will provide a useful background for our subsequent discussion.

Both in his famous paper on truth for formalized languages (1933) and in his subsequent paper on logical consequence (1936/2002), Tarski starts out with an informal, philosophically relevant notion; he then seeks to develop an appropriate formal framework to capture the main features of that notion in a systematic, mathematically tractable way. In the case of truth, the starting point is the correspondence conception of truth, which he claims dates back to Aristotle. In the case of logical consequence, he is somewhat less precise and refers to the 'common' or 'everyday' notion of logical consequence, although the notion of consequence operative in mathematics is also clearly in the background.[3]

These two conceptual starting points allow Tarski to formulate what he describes as *conditions of material adequacy* for the formal accounts, i.e., conditions that any formal treatment of the concepts in question would have to satisfy to count as adequate. (He also formulated criteria of formal correctness, which pertain to the internal exactness of the formal theory.) In the case of truth, the basic condition of material adequacy is the well-known T-schema; in the case of logical consequence, the conditions are the properties of necessary truth-preservation and of validity-preserving schematic substitution.[4]

As Tarski goes on to argue, the formal treatments he proposes (somewhat unsurprisingly) both pass the test of material adequacy he has formulated himself. There is nothing ad hoc about this fact, however, since the conceptual cores of the notions he is after were presumably captured in the corresponding conditions, which can thus serve as conceptual 'guides' for the development of the formal treatments. Indeed, the fact that Tarski can formulate relatively informal but nonetheless *precise* desiderata for his purposes is one of the philosophical strengths of his analyses, both for truth and for logical consequence.

1.2 Carnapian explication

Carnap's first explicit discussion of the notion of explication occurs in the article 'Two concepts of probability' (1945, p. 513), where he introduces the term 'explication' as "an adaptation of the terminology of Kant and Husserl" (a historical reference he repeats in *Logical foundations of probability*).[5] Two years later, in *Meaning and necessity* (1947), he characterizes the corresponding method and goals as follows:

> The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs among the

[3] It has been argued (e.g. Smith 2011) that the idea of a pre-theoretical, intuitive notion of logical consequence is highly problematic. More basically, while logical consequence may be an informal notion, it is already a *theoretical construct* that comes with a long history attached to it. (More on this below.)

[4] In Dutilh Novaes (2012b), a 'conceptual genealogy' is presented in order to unearth the historical origins of these two desiderata for logical consequence.

[5] Sometimes this initial discussion of explication is overlooked (including in Reck 2012).

> most important tasks of logical analysis and logical construction. We call this the task of explicating, or of giving an *explication* for, the earlier concept; this earlier concept, or sometimes the term used for it, is called the *explicandum*; and the new concept, or its term, is called an *explicatum* of the old one. (Carnap 1947, pp. 7–8, original emphasis)[6]

Carnap goes on to propose explications for the concepts of meaning and necessity in this book. In other works, he approaches the concepts of analyticity, confirmation, probability, entropy, etc. along the same general lines (Carnap 1942, 1945, 1950, 1978, etc.). It is important to notice right away that Carnapian explication is a *process*, not merely a relation between explicandum and explicatum. In fact, it is a complex process that involves several distinctive steps. It starts with a preliminary informal clarification of the explicandum, then continues with the elaboration of a suitable explicatum, and culminates in the application of the explicatum to circumstances or situations where the explicandum originally played a central role (a step he calls 'interpretation').

Carnap's discussion of his methodology in *Meaning and necessity* is quite brief. A more elaborate discussion of explication, especially concerning what counts as an adequate formalization in this context, occurs in his next book, *Logical foundations of probability* (1950, Chap. 1). He now introduces explication more succinctly in the following way:

> By an *explication* we understand the transformation of an inexact prescientific concept, the *explicandum*, into an exact concept, the *explicatum*. (Carnap 1950, p. 1, original emphasis)

A few pages later, Carnap formulates four corresponding criteria of adequacy:

> A concept must fulfill the following requirements in order to be an adequate explicatum […]: (1) similarity to the explicandum; (2) exactness; (3) fruitfulness; (4) simplicity. (Ibid., p. 5)

Here exactness and simplicity can be taken to be purely internal criteria, similar to Tarski's criteria of formal correctness. Similarity to the explicandum comes close to what Tarski refers to as 'conditions of material adequacy', namely that the explicatum should reflect the conceptual core of the explicandum in question. But fruitfulness—both the least developed and the most interesting of Carnap's desiderata—seems a true novelty relative to Tarski's discussion, which focuses on formal correctness and material adequacy alone.

In what follows, we discuss the three central desiderata for Carnapian explication further: exactness, similarity, and fruitfulness.[7]

[6] As the reference to 'logical analysis' and 'logical construction' indicates, Carnap views explication as the successor to related methodologies used earlier by Frege, Russell, etc. This includes 'rational reconstruction', as appealed to in his previous work. See Carus (2007) and Beaney (2013) for historical background.

[7] Simplicity seems much less important, both for Carnap's and for present purposes; it will thus be left aside in this paper. (Much of what we will say about exactness should also apply to simplicity, however.)

1.3 Exactness

It is well known that one of the cornerstones of logical empiricism, with Carnap as a central figure, was the idea of emulating the methods of the exact sciences to discuss philosophical issues. Many of the members of the Vienna Circle were themselves accomplished scientists; thus it is not so surprising that they set out to apply these methods more generally, including adhering to the same standards of rigor, exactness, and precision used to judge the mathematically framed theories of, e.g., physics. An immediate question is, however, whether this form of 'scientism' is warranted in the case of subject matters that may be inexact by their very nature. In particular, it is not obvious that philosophical issues can be investigated properly, and without residue, with formal tools.[8]

The general idea of the inadequacy of informal, ordinary languages ('*Sprachen des Lebens*' in Frege's terms) for scientific and philosophical inquiry dates back at least to Frege's *Begriffsschrift* (Frege 1997, pp. 49–51) and runs through the tradition culminating in the Vienna Circle and in Carnap's writings. This inadequacy is often attributed to the perceived lack of exactness afforded by ordinary languages, thus justifying the introduction of more formal expressive means. In particular, concerns about exactness are frequently put forward as concerns about *expressivity*. The basic idea is that informal languages, because of their inexactness, do not allow one to express core concepts or distinctions adequately (such as the mathematical distinction between continuity and uniform continuity).

However, when reading Carnap on explication one often gets the impression that more seems to be at stake regarding exactness. It may even appear that exactness is the main desideratum for the whole enterprise after all, insofar as it is presented as an essential feature of any systematic scientific approach. Thus he writes:

> The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an *exact* form, so as to introduce the explicatum into a well-connected system of scientific concepts. (Carnap 1950, p. 7)

The emphasis on exactness also suggests that Carnap's conception of explication is indeed (as argued in Carus 2007) an instantiation of the ideals of the Enlightenment: "the ambition of shaping individual and social development on the basis of better and more reliable knowledge than the tangled, confused, half-articulate but deeply rooted conceptual systems inherited from our ancestors" (Carus 2007, p. 1).[9] "Better and more reliable knowledge" is quintessentially understood as *scientific* knowledge in this context, and scientific knowledge in turn must adhere to the highest standards of mathematical exactness.

From this point of view, it is quite clear why Carnap insists so much on exactness, to the point that it may appear to be a goal *an sich* for explication. But as the Enlight-

[8] This is a key objection to Carnap's approach in Strawson 1963; cf. Reck 2012. (We will come back to it later.)

[9] For an alternative and less far-reaching account of Carnap's basic motivations, and thus a different interpretation of the primary goals of explication, see Richardson (2013).

enment ideals essentially concern *improving lives* on the basis of scientific knowledge, shouldn't exactness be subordinated to what is arguably the most important requirement for a successful Carnapian explication, namely *fruitfulness*? In other words, it would seem that Carnap's 'pragmatism' should ultimately gain the upper hand over his 'scientism' (as argued in Carus 2007, Ch. 11).[10] Yet in that case, we still need an answer, or at least a more fully worked-out answer, to the question of why exactness is fruitful in this context. Appeals to expressivity, one the one hand, and to improving lives, on the other, only provide partial, preliminary responses.

Sometimes (e.g. in the first passage from *Logical foundations of probability* quoted above) Carnap also seems to suggest that exactness/inexactness are binary, all-or-nothing concepts: the explicandum is inexact, and it is replaced by the exact explicatum. In practice, however (and as Carnap indicates in the passage from *Meaning and necessity* also quoted above), exactness is a comparative concept, not an absolute one, indeed one that comes in degrees. So an explication consists in transforming a *less* exact concept into a *more* exact one, not in transforming something inexact into something exact. If this is the case, it also becomes apparent that explication is a process that can be *iterated*, so that the explicatum of an explication may become the explicandum for another explication, etc. This observation is further supported by the inherently open-ended nature of explication.[11]

The idea of iterated explications is in fact a very useful one with respect to a number of episodes in the history of logic, where antecedent practices or concepts provide the material for regimentations and systematizations by means of logical theorizing. A good example is the development of the concept of a deductive argument, and thus the concept of logical consequence, in ancient Greece. Here the starting point were the dialectical practices of debate in the early Academy, presumably captured in Plato's dialogues. (These in turn were already regimentations of more 'mundane' dialogical practices, not restricted to this niche of specialists.) These practices constitute the explicandum for a process leading to Aristotle's regimentation of a game of questions and answers, as presented in Books I and VIII of the *Topics*. The latter, in turn, led to the crystallization of a more technical, theoretical notion of *syllogismos* (roughly, a deductive argument) for which Aristotle provides (slightly varying) definitions in a number of his writings (*Prior Analytics, Topics, Sophistical Refutations, Rhetoric*), such as:

> A *syllogismos* is an argument (*logos*) in which, certain things being posited, something other than what was laid down results by necessity because these things are so. (*Prior analytics*, 24b18–20)

[10] Even if exactness and fruitfulness are distinct criteria in themselves, if exactness is not to be a goal *an sich*, it should be judged with respect to the pragmatic concerns that seem to motivate the whole enterprise.

[11] In Carus (2007), an implicit "dialectical" conception is attributed to the mature Carnap that underscores and further enriches this point. The corresponding adoption of formal frameworks is described thus: "On the one hand, then, are frameworks of (relatively) precise, hard concepts, on the other hand is the activity of practical decisions among such frameworks. These decisions are at best partly extricable from the entire worlds of practical decisions, which are generally conducted in ordinary, pre-systematic language, i.e., in softer, less precise concepts. They may be hardened up, just as, in the perspective of rational reconstruction, the concepts of ordinary scientific language were progressively upgraded and replaced. But in the new perspective, such progress is no longer a one-way street. The practical realm kicks back." (Carus 2007, p. 21).

This definition of a *syllogismos* then becomes the explicandum for yet another explication. The result is a further explicatum, namely the formal, rigorous system presented in the first chapters of the *Prior Analytics* (the deductive system based on the four forms of categorical propositions, with pairs of premises that do or do not produce conclusions resulting by necessity). Finally, and beyond Aristotle himself now, that system—the one we know as *syllogistic*—can be precisified further once again, e.g. by the formulation of a corresponding deductive system in modern notation (as done by Corcoran, Smiley, and others). Altogether, this constitutes an extended, open-ended chain of explications.

It is clear that, with each iteration of the explication process, we obtain an increase in exactness. But at each step, something is 'transformed' as well, resulting both in gains and in losses. Thus, in the step from the informal dialectical practices to the definition of a *syllogismos*, what is 'lost' is the wide class of arguments that are not necessarily truth-preserving, but are regularly used in these dialectical contexts too (induction, analogy, example). In the step from the definition of a *syllogismos* to the formal system of the *Prior Analytics*, there is a further narrowing of scope: the theory now only covers an even smaller class of arguments, namely those composed exclusively of categorical sentences (All A is B, No A is B, Some A is B, Some A is not B), with two premises and one conclusion.

To sum up this section, it is evident that the criterion of exactness is at the core of Carnapian explication, especially if his project is understood broadly as an Enlightenment project: breaking free from the shackles of a primitive, unsystematic conception of reality, or from entrenched, crippling social practices, in favor of scientific knowledge and arrangements informed by it. Nevertheless, Carnap's pragmatist tendencies should, and do, take pride of place in the end (if Carus' reading is to be trusted). Thus, fruitfulness is ultimately the most significant requirement for an explication overall. But before turning to fruitfulness in detail, let us reconsider another Carnapian desideratum, similarity.

1.4 Similarity

Similarity is a rather weak requirement for Carnap, as he makes clear from the start:

> The explicatum is to be similar to the explicandum in such a way that in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted. (Carnap 1950, p. 7)

Indeed, it is Carnap's apparent denigration of the similarity desideratum that motivates much of the criticism by Strawson (1963) and other 'ordinary language philosophers'. As such critics point out, if the explicandum is inherently inexact and non-rigorous (as a concept belonging to the "conceptual systems inherited from our ancestors", in Carus' words), an explicatum that aspires to be exact will necessarily misrepresent the inexact explicandum, leading to an unfortunate 'change of subject' (Dutilh Novaes and Geerdink forthcoming). On the surface, this misses the point of explication, of course, which is *not* to offer a faithful and accurate account of these 'pre-theoretical concepts' inherited from our ancestors, but rather to *replace* them with more precise concepts *in*

the relevant contexts.[12] One might, however, view such criticism as targeting the very Enlightenment commitments underlying Carnapian explication. But then the debate would have to be shifted to a deeper level: it is no longer a criticism of explication as such, but of the perceived scientism at the core of the Enlightenment, tied to a lack of appreciation for the know-how embedded in many traditional customs and techniques, including ordinary language.

In any case, the degree of similarity in a Carnapian explication will always be less than total, given that its explicatum is intentionally more exact than its explicandum. Nevertheless, similarity remains of some importance in this context, even if in a rather weak sense.[13] This is because there is still the need to justify the claim that a given explicatum is an explicatum *for a given explicandum*, i.e. that they roughly 'talk about' the same phenomena and that they can be used interchangeably, at least in the relevant cases. As this indicates, the issue of similarity in explication (and in formalization more generally) is partly an issue of intentionality, an issue of aboutness. But there is also a pragmatic aspect, concerning the question of whether the explicatum can indeed be used for the same (or at least sufficiently similar) purposes as the explicandum.[14]

A frequently neglected aspect of Carnap's approach is that he is sensitive to this issue, at least to some degree. Thus he recommends a *clarification* of the explicandum before the explication process properly speaking begins. This is to be achieved as follows:

> Even though the terms in question are unsystematic, inexact terms, there are means for reaching a relatively good mutual understanding as to their intended meaning. An indication of the meaning with the help of some examples for its intended use and other examples for uses not now intended can help the understanding. An informal explanation in general terms may be added. All explanations of this kind serve only to make clear what is meant as the explicandum; they do not yet supply an explication, say, a definition of the explicatum; they belong still to he formulation of the problem, not yet to the construction of an answer. (Carnap 1950, p. 4)

Again, the clarification at issue appears to address in part concerns about intentionality, i.e., to ensure that it is clear what we talk about when producing an explication. In fact, in the passage just quoted Carnap seems to suggest that it is largely a matter of inter-

[12] Note that the explicandum does not necessarily become superfluous in non-scientific, everyday contexts, i.e. the explicatum does not replace it completely there; cf. Maher (2007), also again Carus (2007) (as quoted in the previous footnote).

[13] In Carnap's words: "Since the explicandum is more or less vague and certainly more so than the explicatum, it is obvious that we cannot require the correspondence between the two concepts to be a complete coincidence. But one might perhaps think that the explicatum should be as close to or as similar with the explicandum as the latter's vagueness permits. However, it is easily seen that this requirement would be too strong, that the actual procedure of scientists is often not in agreement with it, and for good reasons" (Carnap 1950, p. 5).

[14] As Carnap writes later on: "Explication gives us improved new concepts that can serve the same purposes as the ordinary concepts that created the puzzles; the problems are solved by using the new language instead of the ordinary language in the problematic contexts" (Carnap 1963, paragraph 19). Note that speaking about "the same purposes" here would seem to require a firm grasp of those original purposes.

subjective agreement on the subject matter of an explication. But ultimately more is at stake, namely an adequate understanding of the explicandum and its original/intended uses, which will be fed into the explication. This is a vital component for the success of the enterprise, one that Carnap arguably acknowledges only insufficiently (cf. Reck 2012).

Reconsidering Carnap's methodology in this respect, including its strengths and weaknesses, has something important to offer for a number of recent philosophical debates. Take for example the debate on logical consequence, re-ignited by John Etchemendy's 1990 book, *The concept of logical consequence*. In that book, Etchemendy argues that Tarski's formal analysis of the concept of logical consequence in terms of quantification over models fails because it does not adequately capture the 'intuitive notion' of logical consequence. However, he does not tell us what exactly this intuitive notion is supposed to be, i.e., he does not provide an adequate *clarification* of the concept, even relative to Carnap's fairly relaxed standards (as argued, e.g., in Shapiro 2005). The elusiveness of the so-called 'intuitive notion' then considerably hampered progress in the ensuing debates.

Crucially for us, clarifying the explicandum in an explication is an *informal* activity, as acknowledged by Carnap himself. As such, it corresponds roughly to a 'conceptual analysis', though not in the sense of providing necessary and sufficient conditions. (In his response to Strawson's criticism, Carnap admitted that instead something like Strawsonian "connective analysis" might be useful in such contexts; cf. Reck 2012.) Moreover, there is a related aspect of the explication enterprise that remains inherently informal and inexact: the relation between an explicandum and an explicatum "cannot itself be precise" (Carus 2007, p. 279), given that one of the relata, the explicandum, is an informal notion.[15]

While Carnap accords considerable leeway to an explication with respect to similarity, when can we say that it fails this requirement? This is another aspect addressed only superficially by him. A basic way to measure such failure, suggested by him implicitly, is to compare the respective extensions of the explicandum and the explicatum, i.e. the 'things' (of the appropriate categories) falling under each of them. They should roughly coincide, though some mismatch is allowed (see the Fish vs. Piscis example below). Extensional inadequacy may come in two ways: *overgeneration*—the explicatum's extension contains items not contained in the extension of the explicandum—and *undergeneration*—the explicandum's extension contains items not contained in the extension of the explicatum.

As an example of such inadequacy, we can consider (first-order) Peano Arithmetic. As noted by Carnap himself (1950, p. 17), PA-1 is satisfied not only by the standard sequence of the natural numbers, but also by a multitude of non-standard models. So arguably, PA-1 *overgenerates* with respect to its explicandum, namely 'normal', basic arithmetic. In the other direction, the fact that any formal system containing arithmetic is incomplete may be seen as a case of *undergeneration*, insofar as the system is not able to generate all the truths about the underlying structure. PA1 might still be viewed as an adequate explicatum in various contexts, namely if it proves useful

[15] Smith (2011) presents similar considerations concerning two of the premises of a 'squeezing argument'.

with respect to the particular purposes pursued in them. But of course, this leads us back to fruitfulness.

It should be clear by now that similarity, albeit a weak desideratum, cannot be disregarded entirely. This has significant implications, especially because the desiderata of similarity and fruitfulness are in tension with one another, as we will argue further below.

1.5 Fruitfulness

And now, last but certainly not least, let us turn to fruitfulness, the requirement that Carnap himself seems to view as the most important one. This is what he writes about it:

> The explicatum is to be a fruitful concept […] useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept). (Carnap 1950, p. 7)

Carnap does not say much else about the ways in which an explication should, or could, be fruitful. The idea of an explicatum leading to many universal statements may be understood in terms of predictive power and testability: if it allows for the formulation of such statements, it allows for many predictions that can be tested. In the case of non-empirical theories, allowing for the derivation of many theorems might amounts to the general idea of the 'fruitfulness' of concepts and definitions dear to Frege, among others (Tappenden 1995), although what exactly that amounts to remains to be clarified.

But surely, there must be more to fruitfulness than the formulation or derivation of universal statements. After all, Carnap seems to be after explicata that truly improve an agent's epistemic and pragmatic situation. How else, then, could the notion of a fruitful explication be unpacked? Let us look at his well-known Fish vs. Piscis example:

> When we compare the explicandum Fish with the explicatum Piscis, we see that they do not even approximately coincide […]. What was [the zoologists'] motive for […] artificially constructing the new concept Piscis far remote from any concept in the prescientific language? The reason was that [they] realized the fact that the concept Piscis promised to be much more fruitful than any concept more similar to Fish. A scientific concept is the more fruitful the more it can be brought into connection with other concepts on the basis of observed facts; in other words, the more it can be used for the formulation of laws. (Carnap 1950, p. 6)[16]

Here, besides another appeal to universal laws, we encounter the suggestion of an explicatum's ability to connect with other concepts on the basis of observed facts. The presupposition seems to be that some such connections arise between explicatum and other concepts, which could not arise between explicandum and these other concepts. But this can be pushed one step further: Carnap seems to expect that the explicatum will be able to reveal certain things about the phenomenon in question that the explicandum could not reveal, by means of the newly established connections to other concepts.

[16] It is worth noting that this is an example of explication *not* tantamount to formalization, at least not in the strict sense exemplified by modern formal logic.

In other words, Carnap's view seems to be that an explication is useful or fruitful when it delivers 'results' that could not be delivered otherwise (or with much more difficulty), i.e. with the explicandum alone. What this suggests is a conception of explication as a method for *discovery*, as opposed to a method for testing or justification alone. The goal is to produce *new knowledge* about the phenomena to which the explicandum pertains. In the example above, the concept of Piscis is meant to reveal properties of the animals in question that the concept of Fish would fail to deliver; likewise for other explications.

This observation, if apt, has several important consequences. To begin with, if the explicatum is supposed to add something to our knowledge and understanding of certain phenomena, something that the explicandum alone could not deliver, then a fruitful explication is one where there is some kind of *mismatch* between explicandum and explicatum. The mismatch at issue corresponds precisely to the explicatum's capacity for generating new knowledge about certain phenomena.[17] In this way, explication reveals itself as a cognitive tool leading to discoveries and new insights.

In addition, if what accounts for the fruitfulness of an explication is such a mismatch between explicandum and explicatum, then it becomes clear why fruitfulness is at odds with the requirement of similarity. As is evident now, a less-than-perfect degree of similarity between explicandum and explicatum is not only a tolerable, contingent upshot of explication—it is the *very goal* of a fruitful explication. This in turn implies that, insofar as it is accountable both to similarity and to fruitfulness, Carnapian explication is an inherently paradoxical enterprise. (We will return to the last point in the final section.)

2 Formalisms as cognitive tools

In this section, we continue our discussion of the desideratum of fruitfulness for explication, broadly construed, but now beyond Carnap. Specifically, we turn to empirical findings on human reasoning, in particular on the effects of reasoning based on certain 'formulas': the so-called SPRs, statistical prediction rules. We discuss two examples in this connection: the medical diagnosis of mental illness, and a formula used to calculate the future price of Bordeaux wines. While these are not examples of Carnapian explication per se (though there are similarities or analogies), they illustrate nicely why and how reasoning with formal tools can significantly alter the manner in which a problem is tackled. Our more general conclusion will be two-fold: that the desideratum of exactness should be subsumed to fruitfulness in many contexts; and that the fruitfulness of explication might be understood in cognitive terms, i.e. as affording a cognitive boost to the agent.

2.1 Intuitions versus formulas: SPRs[18]

In his bestseller, *Thinking, fast and slow* (2011), Daniel Kahneman dedicates a whole chapter (aptly titled 'Intuitions vs.formulas') to the question of the differences between

[17] Note that we are conceiving of the mismatch as a certain capacity, not, or at least not directly, in terms of 'concepts' or 'meanings'. (We will come back to this issue at the end of the paper.)

[18] This section is based primarily on Bishop and Trout (2002, 2005) and Kahneman (2011, Chap. 21), including the data cited.

reasoning with 'formulas' and reasoning on the basis of intuitive judgments. The 'formulas' in question, known as SPRs, are simple mathematical algorithms, grounded in prior statistical data, which seem to outperform intuitive human reasoning on a number of practical problems. The seminal work on this topic is *Clinical vs. statistical prediction* (1954), by psychologist Paul Meehl. As Kahneman explains:

> Meehl reviewed the results of 20 studies that had analyzed whether clinical predictions based on the subjective impressions of trained professionals were more accurate than statistical predictions made by combining a few scores or ratings according to a rule. In a typical study, trained counselors predicted the grades of freshmen at the end of the school year. The counselors interviewed each student for forty-five minutes. They also had access to high school grades, several aptitude tests, and a four-page personal statement. The statistical algorithm used only a fraction of this information: high school grades and one aptitude test. Nevertheless, the formula was more accurate than 11 of the 14 counselors. Meehl reported generally similar results across a variety of other forecast outcomes, including violation of parole, success in pilot training, and criminal recidivism. (Kahneman 2011, p. 222)

Since 1954, much more research has been done on the performance of SPRs, and the results are very robust (see Grove and Meehl 1996 for a comprehensive meta-analysis): SPRs outperform intuitive human judgment in about 60 % of cases; and the remaining 40 % show the same degree of accuracy in the two cases, which still means a win for SPRs, as they are normally less costly (cognitively and otherwise) than expert judgment.[19]

One domain in which these statistical formulas tend to be particularly successful is medical diagnosis (Meehl's original topic of research). For example, the Goldberg Rule, which predicts whether a psychiatric patient is neurotic or psychotic on the basis of a Minnesota Multiphasic Personality Inventory (MMPI) profile, is accurate in 70 % of cases, whereas expert judgment's success rate ranges between 55 and 67 % (Bishop and Trout 2002). This may seem counterintuitive at first sight. How can simple formulas outperform recognized experts, with years and years of experience in their fields? But the uncomfortable truth is that, far from being an impartial, purely objective affair, medical diagnosis performed purely clinically (i.e. based on the impressions of the clinician and her expertise) tends to be influenced by a number of apparently irrelevant factors.

This phenomenon can be conceptualized in terms of *cognitive biases* (Croskerry 2003). For example, a doctor who diagnoses disease X in a given patient is significantly more likely to diagnose X in another patient whom she sees on the same day than if there had not been a previous diagnosis of X that day. This tendency can be explained in terms of a phenomenon known as *availability bias*, or 'what comes to mind' bias (Mamede et al. 2010). After having seen a patient apparently suffering from X, the idea of disease X is more vividly present in the doctor's mind, and thus is more likely to occur to her again soon thereafter. Medical diagnosis is also affected by the

[19] See Bishop and Trout (2002, 2005) for further discussions of the astounding success of SPRs.

well-known phenomenon of *confirmation bias*, which leads the clinician to look for evidence that confirms her diagnosis but not for evidence that would disconfirm it. Indeed, there is a variety of such cognitive tendencies that seem to affect medical diagnosis negatively. In contrast, SPRs block, at least to some extent, the effects of these cognitive tendencies in medical diagnosis. (More on this point shortly.)

Another much-discussed example in the SPR literature is the Ashenfelter formula for determining the future quality of wines produced in the Bordeaux region of France (Ashenfelter et al. 1995). In general, high quality wines only peak after many years of maturation, but the lots are usually sold as soon as they are produced. So it is of the utmost importance, both for sellers and buyers, to be able to predict accurately the future price of these wines, based on their estimated future quality. Normally this is done on the basis of current prices of young wines, which are largely determined by the opinions of experts (who swirl, smell, and taste to determine the quality of the wine).

Now, it is conventional wisdom that the best wines are produced when the summer is warm and dry, and when the spring is wet. Based on this observation, the economist and wine lover Orley Ashenfelter developed a simple formula to predict the quality of the vintage for a red Bordeaux wine decades in advance, which takes in only three parameters as values: summer temperatures, rain at harvest time, and total rainfall in the previous winter. This simple formula forecasts future prices much more accurately than predictions based on the current prices of young wines, in fact with more than 90 % accuracy—significantly better than predictions based on expert opinion alone.

2.2 Formulas as cognitive artifacts

How can we make sense of the fact that these statistical formulas so often outperform intuition-based 'human' evaluations and predictions? Based on the cases above, a plausible answer stems from the observation that human intuitive judgment tends to be influenced by external, variable, and often irrelevant factors in general (context-sensitivity).[20] And as already mentioned, some of these tendencies are aptly conceptualized in terms of cognitive biases. Importantly, the relevant formulas are not only efficient predictors: they are also usually quite simple, thus entailing a low level of cognitive demand involved.

Moreover, human judgment is not constant: expert radiologists contradict their own verdict of 'normal' and 'abnormal' upon examining the same chest X-ray on different days in 20 % of the cases. Formulas, by contrast, are constant: a certain input will always yield the same output, no matter the circumstances. Finally, formulas focus on a small number of parameters, which, if well chosen, capture the essence of the causal relations between the phenomena in question, leaving aside irrelevant or less relevant factors. But naturally, the crucial qualification is 'if well chosen' and, more generally, whether the used formulas are suitably designed or not—a poorly engineered formula will of course not perform well.

[20] As Kahneman (2011, p. 225) notes: "The prospects of a convict being granted parole may change significantly during the time that elapses between successive food breaks in the parole judges' schedule."

Typically, SPRs are designed through a 'bottom-up' approach: previous statistical data allow for the identification of certain patterns in the phenomena at issue, and through 'reverse-engineering' one arrives at a formula that fits well with the data.[21] The assumption is that these patterns are constant, and thus, that presently known variables (e.g. summer temperature and rainfall) will allow us to make reliable predictions of the values sought (e.g. future price of young Bordeaux wines). The accuracy of the reverse-engineered formulas can then be further put to the test, so that those that perform poorly in terms of predictions are either abandoned or revised. This flexibility and revisability is another strength of the approach: it is relatively easy to change the formula one uses, but it is hard to change entrenched intuition-based reasoning patterns (especially given the well-documented phenomenon of overconfidence).

Let us return to cognitive biases once more, as related to the issue of formalization. In Dutilh Novaes (2012a), the thesis defended was that formal languages and formalisms more generally are best understood as cognitive technologies or tools that can be used to assist human reasoning. At first glance, this claim is not particularly controversial: naturally, it is easier to make calculations, to draw inferences, etc. with pen and paper than mentally; but in principle, most calculations, inferences, and the like can be performed without pen and paper as well. However, the additional, more far-reaching claim is that formalisms not only augment human 'brain power': they literally *transform* how we reason.

Crucially, formalisms can compensate for a number of cognitive tendencies that are disadvantageous in certain circumstances (e.g. medical diagnosis), thus leading to sub-optimal judgments. In the cases mentioned above, the main point is that SPRs seem to help human reasoners to 'stick to the essentials' rather than being distracted by irrelevant or less relevant (contextual) factors. And this is precisely due to the fact that these formulas are tied to purely mechanical procedures, which means that they require no 'insight or ingenuity' to be applied (arguably, the main source of interference from irrelevant external factors), thus countering many of the cognitive biases discussed in the literature.[22]

2.3 SPRs, Carnapian explication, and cognitive fruitfulness

At this point, the reader may wonder what the relevance of SPRs is for the main topic of this paper, namely Carnap's notion of explication. Are we saying that SPRs simply *are* instances of Carnapian explication? This would be taking the argumentation a step too far, even though there are a number of interesting similarities or analogies between the two sides. In the case of the Ashenfelter formula, for example, we may see the informal idea that the best wines are produced when the summer is warm and dry and

[21] As Bishop and Trout put this point: "To complete the proper linear model, we need a reasonably large set of data showing how these cues correlate with the target property (the market price of mature Bordeaux wines). Weights are then chosen so as to best fit the data: they optimize the relationship between P (the weighted sum of the cues) and the target property. An actuarial model along these lines has been developed (Ashenfelter et al. 1995)" (Bishop and Trout 2002, p. 198).

[22] A more detailed exposition of why the 'mechanical' manipulation of formalism has the effect of countering cognitive biases can be found in Dutilh Novaes (2012a).

when the spring is wet as constituting, or at least as corresponding to, an explicandum, and the formula itself as the respective explicatum. Moreover, the fact that the formula is arrived at on the basis of statistical analysis is perfectly in the spirit of Carnapian explication. However, in other cases it is harder to uphold the analogy, and some more general differences remain. In particular, SPRs are not primarily meant to be cases of concept formation, as is Carnap's goal, though this may sometimes happen too (e.g., with respect to medical categories).

A more fundamental point is this: SPRs display two core characteristics attributed to explications by Carnap—exactness and fruitfulness.[23] Similarity is less operative. Because SPRs typically disregard much of the information that agents tend to consider relevant for making a judgment, it might even seem that SPRs systematically violate the similarity requirement (that is, if the intuitive approach of an expert is seen as the explicandum). Then again, given how much Carnap himself deemphasizes similarity for explication, this does not constitute a strong objection to comparing SPRs to Carnapian explicata. Note also that general concerns of 'aboutness', or about satisfying the same purposes, typically do not arise here, as SPRs are designed to provide answers to the same (or similar enough) questions posed to the experts: Is this patient best diagnosed as psychotic or neurotic? What is the estimated price of this vintage of Bordeaux in 20 years from now?

But our two main reasons for considering SPRs in a discussion of Carnapian explanation are the following: (1) They instantiate the Enlightenment idea of relying on (to quote Carus again) "better and more reliable knowledge [modes of reasoning] than the tangled, confused, half-articulate but deeply rooted [modes of reasoning belonging to the] conceptual systems inherited from our ancestors." (2) They provide a vivid illustration of how exactness can be fruitful, as SPRs represent real improvement over the results of 'less exact', more intuitive modes of reasoning. Exactness can thus be subordinated to fruitfulness in such cases, and this from a cognitive perspective.[24] Let us now elaborate on each of these two points in turn.

(1) The concept of 'debiasing', understood (as in Dutilh Novaes 2012a) as the process of compensating for certain cognitive tendencies that are not advantageous in some, or even many, contexts, has an obvious Enlightenment bent to it. Similarly, when Bishop and Trout (2005) defend what they call Ameliorative Psychology, with SPRs as one of their strongest exhibits, it is again an Enlightenment-inspired project they are defending. Now, the Enlightenment nature of Carnapian explication is most conspicuous in his emphasis on exactness, especially if one understands it not only in formal, mathematical terms, but as a property of *scientifically* robust concepts in a more inclusive sense. Indeed, SPRs are scientifically robust in two ways: they are mathematically framed; but they are also robustly grounded in empirical data, given the 'bottom-up' approach adopted to generate

[23] SPRs also satisfy the desideratum of simplicity well. But again, we do not focus on simplicity in this paper.

[24] Likewise for simplicity; indeed, SPRs illustrate nicely how exactness and simplicity can combine fruitfully.

them.[25] Hence, the debiasing effect of SPRs vividly illustrates Carnap's idea of improving one's overall epistemic and pragmatic stance by minimizing the influence of 'biased' modes of thinking (while not disregarding that, in many situations, the 'older' modes of thinking still do just fine and may even be irreplaceable). As this brings to the fore, Carnapian explication is essentially an *ameliorative* project.[26]

(2) Critics of Carnapian explication often focus on (the overemphasis on) mathematical exactness as its most problematic aspect. And it is true that exactness *an sich* will not necessary be illuminating or helpful; in fact, it may have the opposite effect, i.e., that of obscuring what was initially transparent. More generally, it is not clear how one could make a compelling case for the desirability of exactness as a goal in and of itself. In that sense, the critics have a point when they highlight the 'scientism' built into the approach. Now, the most promising strategy to reply to this challenge from a Carnapian point of view seems to be to *subordinate the desideratum of exactness to that of fruitfulness*, i.e., to argue that exactness can be *useful* in various circumstances. Indeed, the conception of formalisms as cognitive tools is based on the recognition of the enhancement afforded by formalisms over informal modes of reasoning. In particular, it is precisely their exactness that gives rise to the debiasing effect of certain formalisms, insofar as implementing largely 'mechanical' procedures has the effect of 'turning off' more familiar patterns of reasoning (as detailed further in Dutilh Novaes 2012a, Chaps. 5–7). This may lead to more reliable results, as the case of SPRs illustrates.

3 The paradox of adequate formalization

Nonetheless, there is a permanent dark cloud disrupting the blue skies of formalization. On the conception just presented, the project of formalization is inherently paradoxical, in its Carnapian version as well as more generally. On the one hand, a particular formalization has to be sufficiently similar to its target phenomenon to be rightly described as a formalization *of that target phenomenon*, and also to be applicable to the *same, or at least closely related, purposes*. On the other hand, the formalization will be more useful insofar as it says something about the target phenomenon which prior, informal conceptualizations of it *did not reveal*. In other words, an adequate formalization is one that is faithful to the target phenomenon *and* reveals something new about it; there is an obvious tension between these two desiderata. We call this the *paradox of adequate formalization*.

This paradox embodies, indeed, much of the criticism voiced against formal approaches to philosophical questions generally: such approaches may fail to be illuminating, not adding anything new to our pre-existing knowledge of the subject-matter (thus, not informative); or they may 'change the subject', i.e., misrepresent or distort

[25] In most of this paper, we understand Carnap's desideratum of exactness in the narrower (formal or mathematical) sense; but at some points the more inclusive sense also plays a role.

[26] Perhaps surprisingly, Carnapian explication can be viewed an ameliorative project very much in the spirit of Haslanger's (2012) project of social critique, which is also described as ameliorative.

the subject-matter by imposing exactness where there is none to be found (thus, not similar). Critics of Carnapian explication, e.g. Strawson, have identified the same basic tension between exactness and similarity for it (an exact explication of an inexact explicandum will fail the similarity requirement), or alternatively, between exactness and fruitfulness (exactness may well turn what is accessible in informal guise into something obscure, including by neglecting pre-formal aspects of how our purposes were served originally). Both concerns have already been addressed in this paper, including the most promising Carnapian response to them.

But so far critics have not highlighted sufficiently what is arguably the most fundamental tension in Carnap's approach: that between the desiderata of similarity and fruitfulness, which also pull in opposite directions. Thus, the similarity requirement recommends that explicandum and explicatum be 'the same' in some relevant sense, or at least similar enough; in other words, there is pressure for the two to *match.* The fruitfulness requirement, on the other hand, recommends that the explicatum offer something that the prior, informal grasp of the phenomenon (the explicandum) cannot offer so that the explication can be informative; in that sense, there is pressure for the two to *mismatch*.

While this conflict between similarity and fruitfulness in explication is an instance of the paradox of adequate formalization, as we called it, the basic phenomenon pertains not only to formalizations: it pertains to analysis more generally and, as such, is known as the *paradox of analysis* (Beaney 2014). This paradox can be formulated in a variety of forms, some subtler than others. In its simplest form, the two premises of the paradox are: for an analysis to be correct, the analysans must be identical to the analysandum; but for the same analysis to be informative, the analysans must be somehow different from the analysandum. The conclusion is that no analysis can be both correct and informative: if an analysis is correct, it is not informative; if it is informative, it is not correct.

The first use of the phrase 'paradox of analysis' appears to be in C.H. Langford's "The notion of analysis in Moore's philosophy" (1942a), in the *Living philosophers* volume devoted to G.E. Moore (Schilpp 1942). (A reply by Moore is contained in the same volume.) A number of responses to this paradox have been proposed since then. One may of course bite the bullet and accept the conclusion, unpalatable as that may seem: an analysis cannot be both informative and correct. But one may also question one of the premises, or at least specific formulations of the premises that give rise to the paradox. From what has already been said, it is clear what the Carnapian response is: it consists in the rejection of the premise that, for an analysis to be 'correct', analysans and analysandum must be identical.

Indeed, Carnap knew about the Langford–Moore exchange (Carnap 1947, p. 63). Thus, his introduction of the notion of explication, in "Two concepts of probability" (1945), and its elaboration in subsequent writings can be seen as a direct response to this problem as raised by Langford.[27] More specifically, the fact that Carnap not only speaks of 'similarity' instead of 'correctness' (as in the original formulation of the paradox of analysis) but also downplays the latter by making 'fruitfulness' the main

[27] This point was suggested to us by an anonymous referee.

criterion of adequacy, constitutes arguably his attempt to solve, or bypass, the paradox of analysis. Then again, this move is part of a more general shift towards pragmatist ideas in Carnap's later writings as well.

In this sense, a Carnapian should accept the 'mismatch' horn of the dilemma above, which makes good sense from the point of view of explication as an ameliorative project. Again, fruitfulness is valued over similarity, and although Carnap himself did not elaborate much on what he means by fruitfulness, Sect. 2 of the present paper offered an attempt to spell out what it might mean from the perspective of formalisms as cognitive tools. It bears repeating, however, that even a Carnapian cannot ignore similarity altogether, since it is what ensures the right kind of intentionality and purpose-sensitivity for an explication (as noted, Carnap himself did not give it up entirely). Put in Tarski's terminology, something like 'conditions of material adequacy' remains crucial for any formalization project, even if along Carnapian lines this has become part of a more multidimensional procedure. And insofar as this is the case, the 'paradox' is still relevant.[28]

We would like to add one more remark about what is at issue philosophically here. As we saw, Carnap adopts 'similarity' as a desideratum for explication, not 'correctness'. He does not spell out very clearly what similarity amounts to, except for rejecting anything as strong as identity of 'concepts' or 'meanings'. Nevertheless, some weaker form of 'faithfulness' or 'matching' remains relevant, as we pointed out repeatedly. But then, there also remains the question of how best to think about that relation—in extensional terms (as suggested by some of Carnap's remarks and discussed briefly above), in intensional terms (via some coarse-grained notion of 'similarity of intensions'), or perhaps, in terms of some kind of 'continuity of purposes'. The latter might look most promising from a pragmatist perspective; but it is also the option least explored by philosophers so far, including beyond Carnap. Thus there remains clarificatory work to be done on this issue.

Whatever a good answer to this remaining philosophical question might be, examples of formal tools that improve human reasoning in concrete, demonstrable ways, such as SPRs, suggest the following: despite its paradoxical nature, formalization remains a powerful tool for human reasoners. In particular, it can have the welcome effect of countering reasoning tendencies in humans that lead to sub-optimal results in various circumstances. Thus, our overall conclusion is: Yes, formalization inherits the paradoxical nature of its ancestor, namely conceptual analysis, even if not in its crudest form; and in the case of Carnapian explication (and as we noted, formalization by logical means was for Carnap the quintessential form of explication), this translates into the tension between similarity and fruitfulness. But this is not to be viewed as a sign that the enterprise of formalization should be given up altogether; it only suggests that careful methodological reflection is required. At its best, formalization offers a cognitive boost that we, as reasoners, had better make good use of.

[28] A less 'dramatic' way to think about this aspect is in terms of trade-offs between different criteria, rather than attributing an inherent paradoxical nature to formalization and Carnapian explication. (We owe this point to an anonymous referee.) But given our contention that similarity remains significant, both for Carnap and for formalization in general, we stick to our claim that something stronger is involved than mere trade-offs.

References

Ashenfelter, O., Ashmore, D., & Lalonde, R. (1995). Bordeaux wine vintage quality and the weather. *Chance*, *8*, 7–14.

Awodey, S., & Reck, E. H. (2002). Completeness and categoricity, part I: 19th century axiomatics to 20th century metalogic. *History and Philosophy of Logic*, *23*, 1–30.

Beaney, M. (2013). Analytic philosophy and history of philosophy: The development of the idea of rational reconstruction. In E. H. Reck (Ed.), *The historical turn in analytic philosophy* (pp. 231–260). Basingstoke: Palgrave Macmillan.

Beaney, M. (2014). Analysis. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/entries/analysis/.

Bishop, M. A., & Trout, J. D. (2002). 50 years of successful predictive modeling should be enough: Lessons for philosophy of science. *Philosophy of Science*, *69*, 197–208.

Bishop, M. A., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment*. Oxford: Oxford University Press.

Carnap, R. (1942). *Introduction to semantics*. Cambridge, MA: Harvard University Press.

Carnap, R. (1945). Two concepts of probability: The problem of probability. *Philosophy and Phenomenological Research*, *5*, 513–532.

Carnap, R. (1947). *Meaning and necessity*. Chicago: University of Chicago Press.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.

Carnap, R. (1963). Replies and systematic expositions. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 859–1013). Chicago: Open Court.

Carnap, R. (1978). In A. Shimony (Ed.), *Two essays on entropy*. Berkeley: University of California Press.

Carus, A. W. (2007). *Carnap and twentieth-century thought: Explication as enlightenment*. Cambridge: Cambridge University Press.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, *78*, 775–780.

Dutilh Novaes, C. (2012a). *Formal languages in logic*. Cambridge: Cambridge University Press.

Dutilh Novaes, C. (2012b). Medieval theories of consequence. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/entries/consequence-medieval/.

Dutilh Novaes, C., & Geerdink, L. (forthcoming). The dissonant origins of analytic philosophy: Common sense in philosophical methodology. In S. Lapointe & C. Pincock (Eds.), *Innovations in the history of analytical philosophy*. Basingstoke: Palgrave Macmillan.

Frege, G. (1997). In M. Beaney (Ed.), *The Frege reader*. Oxford: Basil Blackwell.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323.

Haslanger, S. (2012). *Resisting reality*. Oxford: Oxford University Press.

Justus, J. (2012). Carnap on concept determination: Methodology for philosophy of science. *European Journal for Philosophy of Science*, *2*, 161–179.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Strauss & Giroux.

Kuipers, T. (2007). Introduction: Explication in philosophy of science. In T. Kuipers (Ed.), *General philosophy of science: Focal issues* (pp. vii–xxiv). North Holland: Elsevier.

Langford, C. H. (1942). The notion of analysis in Moore's philosophy. In P. A. Schilpp (Ed.), *The philosophy of G.E. Moore* (pp. 319–342). LaSalle, IL: Open Court.

Maher, P. (2007). Explication defended. *Studia Logica*, *86*, 331–341.

Mamede, S., van Gog, T., van den Berge, K., Rikers, R., van Saase, J., van Guldener, C., et al. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA: The Journal of the American Medical Association*, *304*(11), 1198–1203.

Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Reck, E. H. (2012). Carnapian explication: A case study and critique. In P. Wagner (Ed.), *Carnap's ideal of explication and naturalism* (pp. 96–116). London: Palgrave Macmillan.

Richardson, A. (2013). Taking the measure of Carnap's philosophical engineering. In E. H. Reck (Ed.), *The historical turn in analytic philosophy* (pp. 60–77). Basingstoke: Palgrave Macmillan.

Schilpp, P. A. (Ed.). (1942). *The philosophy of G.E. Moore*. Chicago: Open Court.

Schilpp, P. A. (Ed.). (1963). *The philosophy of Rudolf Carnap*. Chicago: Open Court.

Shapiro, S. (2005). Logical consequence, proof theory, and model theory. In S. Shapiro (Ed.), *The Oxford handbook of the philosophy of mathematics and logic* (pp. 651–670). New York: Oxford University Press.

Smith, P. (2011). Squeezing arguments. *Analysis*, *71*, 22–30.

Strawson, P. F. (1963). Carnap's views on constructed systems versus natural languages in analytic philosophy. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 503–518). Chicago: Open Court.

Tappenden, J. (1995). Extending knowledge and 'fruitful concepts': Fregean themes in the foundations of mathematics. *Noûs*, *29*, 427–467.

Tarski, A. (1933). *Der Wahrheitsbegriff in den formalisierten Sprachen*. English translation, The concept of truth in formalized languages. In A. Tarski (Ed.). (1956). *Logic, semantics: Metamathematics* (pp. 152–278). Oxford: Oxford University Press. Second edition, 1983.

Tarski, A. (1936). *Über den Begriff der logischen Folgerung*. English translation, On the concept of logical consequence. In A. Tarski (Ed.). (1956). *Logic, semantics: Metamathematics* (pp. 409–420). Oxford: Oxford University Press. Second edition, 1983.

Tarski, A. (2002). On the concept of following logically. *History and Philosophy of Logic*, *23*(3), 155–196.